Assessing the difficulty of the text input task for minority languages

Hugh Paterson III

Hello World!

Cite as: Hugh J. Paterson III. 2015. Assessing the difficulty of the text input task for minority languages. Paper presented at the 4th International Conference on Language Documentation and Conservation (ICLDC).
Link: http://hdl.handle.net/10125/25318

Abstract

Currently no framework exists to evaluate or rank the complexity of the text input task on a per orthography bases. We present on the challenges which must be addressed by a cross-language text input assessment framework. We discuss relevant user experience (UX) considerations for keyboard layouts and unique actions undertaken in the communicative act of 'entextualizing' language (typing). We follow previous work which focuses on majority language text input methods (Bellman & MacKenzie 1998, Castellucci & MacKenzie 2013, MacKenzie 1992, 2002, 2007, MacKenzie & Soukoreff 2002, Soukoreff & MacKenzie 2001, 2003a, b) and apply considerations for minority language orthographies - especially those orthographies which overtly mark tone and other distinctions via diacritics.

The ability to communicate with electronic text based devices is important in this era of globalization. Many minority language users often find it difficult to type in their languages because of the way that orthography/language specific characters are accessed through existing keyboard layouts ([Author] 2012). The keyboard layout is an essential component in text input both on mobile touch screen and traditional devices. Barriers to efficiently using text in digital mediums has a wide impact on language vitality, by affecting the way that language users perceive their language's viability in the 21st century context. The text input challenge has been often acknowledged by minority language users (Esizmetor 2009: 13, Zheltov 2005). Perceptions about the need to be able to use text based digital communication devices has sufficiently challenged language communities leading some to change their orthographies. Simons and Lewis (2010) describe the social practice of literacy (EGIDS levels four and five) as a sign of a healthy language. A text input device which does not intuitively work for language users can be seen as discriminating and be a reason for speakers to choose to not use their language in digital mediums (Trosterud 2012). We propose a language agnostic framework for text input analysis for the benefit of language development efforts and software developers alike.

References:

- Bellman, Tom & I. Scott MacKenzie. 1998. A Probabilistic Character Layout Strategy for Mobile Text Entry. Proceedings of Graphics Interface '98, 168-76. Toronto: Canadian Information Processing Society.
- Castellucci, Steven J. & I. Scott MacKenzie. 2013. Gathering Text Entry Metrics on Android Devices. Proceedings of the International Conference on Multimedia and Human-Computer Interaction - MHCI 2013, 120.1-.8. Ottawa, Canada: International ASET, Inc.
- Esizmetor, David Oshorenoya. 2009. What Orthography for Naijá? Paper presented at Conference on Nigerian Pidgin, University of Ibadan, Nigeria.
- Lewis, M. Paul & Gary F. Simons. 2010. Assessing endangerment: Expanding Fishman's GIDS. Revue Roumaine de Linguistique 55.2: 103–20.
- MacKenzie, I. Scott. 1992. Fitts' law as a research and design tool in human-computer interaction. Human-Computer Interaction 7, 91-139.
- MacKenzie, I. Scott. 2002. Introduction to this special issue on text entry for mobile computing. Human-Computer Interaction 17.2-3: 141-5.
- MacKenzie, I. Scott. 2007. Evaluation of text entry techniques. In I. Scott MacKenzie & Kumiko Tanaka-Ishii (eds.), Text entry systems: Mobility, accessibility, universality, 75-101. San Francisco, CA: Morgan Kaufmann.
- MacKenzie, I. Scott & R. William Soukoreff. 2002. A Character-level Error Analysis Technique for Evaluating Text Entry Methods. A character-level error analysis technique for evaluating text entry methods. Proceedings of the Second Nordic Conference on Human-Computer Interaction -- NordiCHI 2002, 241-4. New York: ACM.
- Paterson, Hugh J, III. 2012. Keyboard layout as part of language documentation: the case of the Me'phaa and Chinantec keyboards. Paper presented at CRASSH Language Endangerment: Methodologies and New Challenges, Cambridge, UK.
- Soukoreff, R. William & I. Scott MacKenzie. 2001. Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic. Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI 2001, 319-20. New York: ACM.
- Soukoreff, R. William & I. Scott MacKenzie. 2003a. Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. Paper presented at Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA.
- Soukoreff, R. William & I. Scott MacKenzie. 2003b. Input-based Language Modeling in the Design of High Performance Text Input Techniques. Proceedings of Graphics Interface 2003 (CIPS, Canadian Human-Computer Communication Society), 89-96. Halifax, Nova Scotia: A K Peters.
- Trosterud, Trond. 2012. A restricted freedom of choice: Linguistic diversity in the digital landscape. Nordlyd (Tromsø University Working Papers on Language and Linguistics) 39.2: 89-104.
- Zheltov, Pavel V. 2005. Minority languages and computerization. The situation in the Russian Federation. OGMIOS 3.3: 8-11.

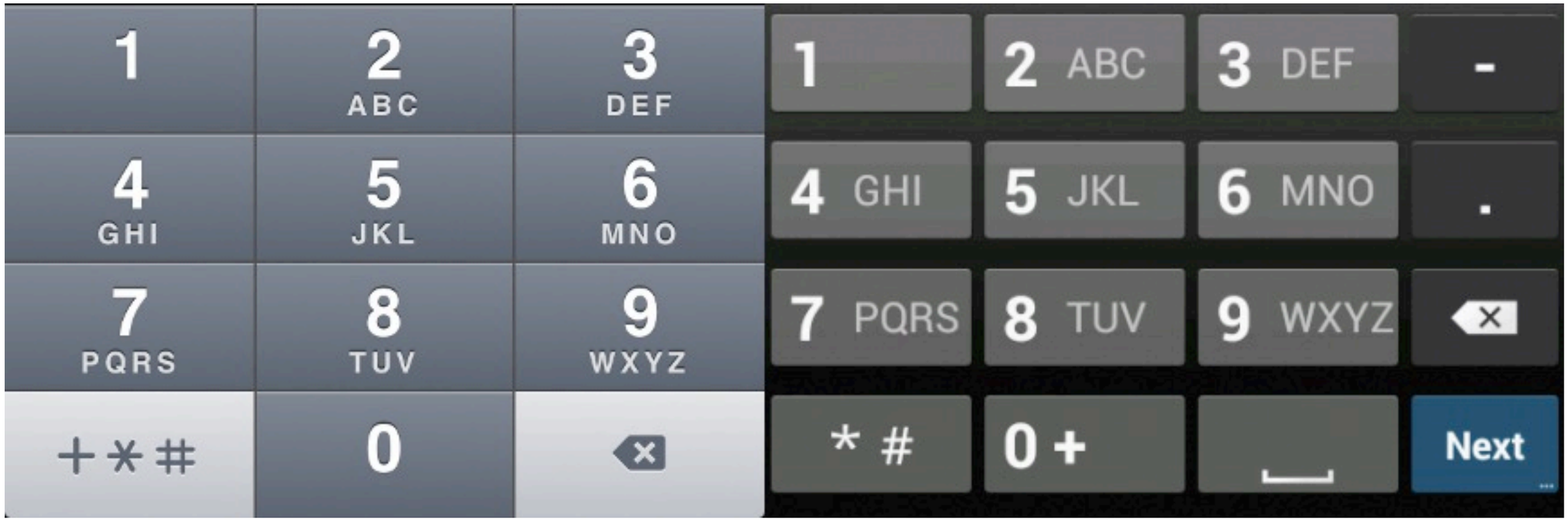
Text input

Basically I mean typing!

- Is not orthography
- Is complex
- Happens in a variety of *environments* - various combinations of *devices* and *sociolinguistic and socio-technological settings*.

Typing happens in different places

With different tools.



Single Finger Keyboard

Multi Character Keyboard

Text input

- And then we want the text to look like different things...

Sample Me'phaa Text

A̱ ngu̱in', tsáan' ninimba̱'la' ju̱ya̱á Jesús, ga̱ju̱ma̱'la' rí phú gagi juwala' ído̱ rí nanújngala' awúun mba'a inii gajmá. Numuu ndu̱ya̱á mála' rí ído̱ rí na̱'nga̱'la' inuu gajmá, nasngájma ne̱ rí gakon rí jañii a̱kia̱n'la' ju̱ya̱á Ana̱'ló', jamí na'ne ne̱ rí ma̱wajún gúkuála'. Indo̱ó má' gí'maa rí ma̱wajún gúkuála' xúgíí mbi'i, kajngó ma̱jráan'la' jamí ma̱'ne rí jañii a̱kia̱n'la', asndo rí náxá'yóo nitháan rí ja'yoo ma̱nindxa̱'la'. [Iyii' rí ni'tháán Santiágo 1:2-4]

Sample Sochiapam Chinantec Text

Hnoh² reh², ma³hiún¹³ hnoh² honh² lɨ³ua³ cáun² hi³ quiunh³² náh², quí¹ la³ cun³ hi³ má²ca³lɨ³ ñíh¹ hnoh² jáun² hi³ tɨ³ jlánh¹ bíh¹ re² ɬ²tɨn² tsú² hi³ jmu³ juenh² tsɨ́³, mɨ́¹juáh³ zia³² hi³ cá² lau²³ ca³tɨ²¹ hi³ taunh³² tsú² jáun² ta²¹. Hi³ jáun² né³, chá¹ hnoh² cáun² honh², hi³ jáun² ɬ¹³ lɨ³tɨn² hnoh² re² hi³ jmúh¹³ náh² juenh² honh², hi³ jáun² hnoh² ɬ¹³ ɬn³ náh² tsá² má²hún¹ tsɨ́³, tsá² má²ca³hiá² ca³táunh³ ca³la³ tán¹ hián² cu³tí³, la³ cun³ tsá² tiá² hi³ lɨ³hniauh²³ hí¹ cáun² ñí¹con² yáh³. [Jacobo Jmu² Cáun² Sí² Hi³ Ca³tɨn¹ Tsá² *Judíos, Tsá² Má²tiáunh¹ Ñí¹ Hliáun³ 1:2-4]

So our fingers dance different dances

Me'phaa

Sample Text

So our fingers dance different dances

Full Text

Me'phaa Sample Text

But dancing is work... right?

Sochiapam
Chinantec
Sample Text

But dancing is work... right?

Full Text

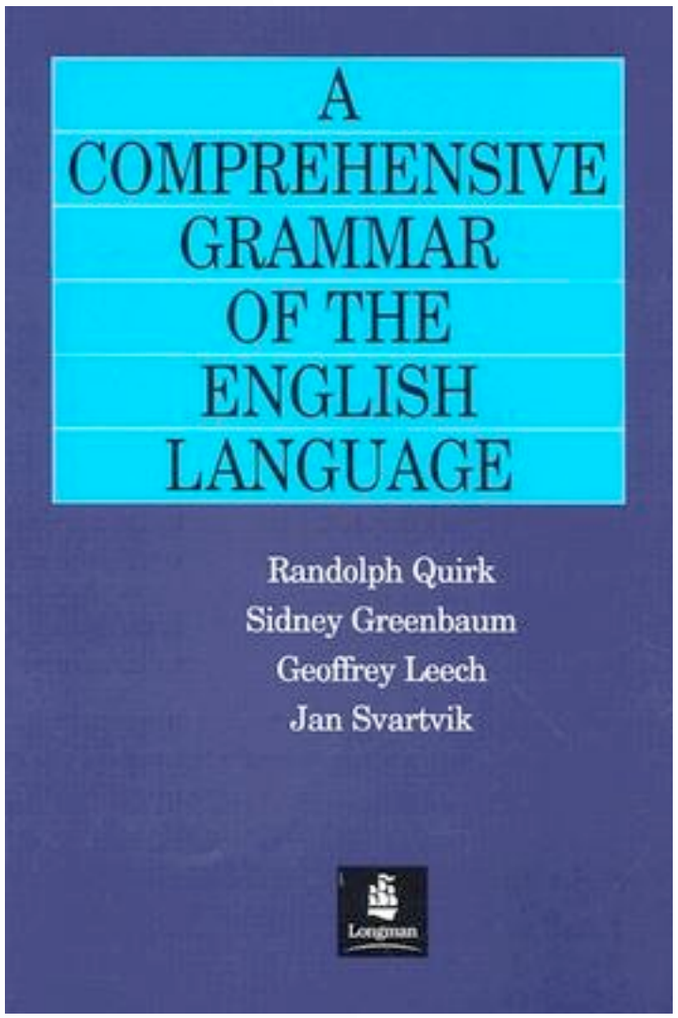
Sochiapam Chinantec Sample Text

Hit load, hand balancing, and finger balancing

Text input

And we hope to do different things by using text...

Filling various social communicative functions



Assumptions about success

- allow the digital text input of an orthography
- allow typing a text without fatigue
- maximize typing speed
- reduce the number of typing errors
- allow rapid mastery of the touch typing method

Conceptualizing the problem space

- Distance - Time equation
  - The shortest distance between two points is a straight line.

Building on what?

- Model 1: Counting chorded keystrokes as a single keystroke (Constable 2001)
  - Does every reach of the fingers count or do only the produced characters count?
- Model 2: Not counting shifted characters at all (Shieh & Lin 1999)
  - Does every character count including punctuation?

Building on what?

- Model 3: Single language/orthography - ENGLISH with assumptions like M = N, M > N, but not on the 10 key layout (Yin & Su 2011)
  - Do we find ourselves in monolingual communicative contexts or in multilingual contexts?
  - Do we find ourselves in contexts where we have more characters than keys?
- Model 4: Not defining a 'character' Mackenzie et al. (in general) -
  - easy: o
  - Difficult: ó, ò, ō, o̩ / o̩, ó̩ / ó̩ , ò̩ / ò̩

Why do we need a comparative framework?

- Can we just change the orthography to fit the keyboard?
  - (Boerger 2007: 133: South Pacific, Cooper 2005: 149, 160: Central Asia, Jany 2010: Americas) vs. (Bailey 2007)
- We know that typing is hard, but how hard (or what is hard about it)?
  - Practically speaking minority language users come to us with a "problem" but are we addressing it with the appropriate solution? How much investigative effort do we put into the text input side of the "literature development" problem space?
- What do we compare?

What to compare

Six criteria

- tapping load distribution
- number of keystrokes
- hand alternation
- finger alternation
- finger posture
- hit direction (little finger to thumb)

Keyboarding Typology

- Single Character Key example - QWERTY
- Single finger keyboard
- Multiple Character Keyboard - T9 phone
- Multiple finger Keyboard

What else is helpful to compare?

- Perceptual distance
- Measuring dissonance

Dissonance

- What does the user see and have to ignore so that they can achieve the results they desire?

Perceptual distance

- The distance the user casually thinks they will travel (or the time taken) to achieve the result

References

Bailey, Dwayne. 2007. Creating a single South African keyboard layout to promote language. *Lexikos* 17.1: 212-25. http://www.ajol.info/index.php/lex/article/view/51533

Boerger, Brenda H. 2007. Natqgu Literacy: Capturing Three Domains for Written Language Use. Language Documentation & Conservation 1.2: 126–53.

Constable, Peter G. 2001. Understanding characters, keystrokes, codepoints and glyphs: encoding and working with multilingual text. In Melinda Lyons (ed.), *Implementing Writing Systems: an Introduction*, 9-33. Dallas: SIL International.

Cooper, Gregory. 2005. Issues in the Development of a Writing System for the Kalasha Language. Ph.D dissertation, Macquarie University.

Eggers, Jan, Dominique Feillet, Steffen Kehl, Marc Oliver Wagner & Bernard Yannou. 2003. Optimization of the keyboard arrangement problem using an Ant Colony algorithm. *European Journal of Operational Research* 148.3: 672-86. http://www.sciencedirect.com/science/article/pii/S0377221702004897

Fauconnier, Gilles. 1985. Mental spaces: aspects of meaning construction in natural language. Cambridge, Mass.: MIT Press (Ch1 pp 3-34).

Ferguson, Charles Albert. 1968. Language development. In A. Fishman Joshua, Charles Albert Ferguson & Jyotirindra Das Gupta (eds.), *Language problems of developing nations*, 27-35. New York: John Wiley & Sons, Inc.

Fishman, Joshua A. 1964. Language Maintenance and Language Shift as a Field of Inquiry. A Definition of The Field and Suggestions for its Further Development. Linguistics 2.9: 32. http://www.degruyter.com/view/j/ling.1964.2.issue-9/ling.1964.2.9.32/ling.1964.2.9.32.xml

Jany, Carmen. 2010. Orthography Design for Chuxnabán Mixe. Language Documentation & Conservation 4.1: 231-53.

Yin, Peng-Yeng & En-Ping Su. 2011. Cyber Swarm optimization for general keyboard arrangement problem. *International Journal of Industrial Ergonomics* 41.1: 43-52. http://www.sciencedirect.com/science/article/pii/S0169814110001058

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

Shieh, Kong-King & Chin-Chiuan Lin. 1999. A Quantitative Model For Designing Keyboard Layout. *Perceptual and Motor Skills* 88.1: 113-25. <Accessed: 2014/06/04>. http://dx.doi.org/10.2466/pms.1999.88.1.113

Spanish
Sample Text

Spanish
Full Text

Spanish
Sample Text

English -
QWERTY
Sample Text

English -
QWERTY
Full Text

English -
QWERTY
Sample Text